What is claimed is:

1.   A semi-automatic construction method for knowledge base of an encyclopedia question answering system, the method comprising the steps of:

(a) designing structure of the knowledge base with a plurality of templates for each entry and a plurality attributes related to each of the templates;

(b) extracting structured information including the entry, an attribute name and attribute values from summary information of the encyclopedia;

(c) extracting unstructured information including an attribute name and attribute values of the entry from a body of the encyclopedia; and

(d) storing the structured information and the unstructured information in corresponding template and attribute of the knowledge base according to the entry.

2.   The method of claim 1, wherein in the step (a), the structure of the knowledge base is constructed with common attribute templates of a common attribute shared in categories of the encyclopedia and individual attribute templates of a specific attribute of an individual category of the encyclopedia, for each entry.

3.   The method of claim 1, wherein in the step (a), attributes having similar meaning are managed as a representative attribute integrally and detail meanings of the attributes are grouped and defined in separate subgroup fields.

4.   The method of claim 1, wherein the step (b) comprises the steps of:

(b-1) recognizing a patterned format of the summary information; and

(b-2) extracting the entry, the attribute name and the attribute values.

5.   The method of claim 4, wherein the step (b-2) comprises the steps of:

(b-2-i) extracting the entry and the attribute name through the patterned format of the summary information;

(b-2-ii) ascertaining whether the attribute name belongs to a valid attribute in an attribute list of the templates of the knowledge base; and

(b-2-iii) only if the attribute name belongs to the valid attribute in the attribute list of the templates of the knowledge base, extracting the corresponding attribute value.

6.   The method of claim 4, wherein the step (b) further comprises the steps of:

(b-3) if the extracted attribute name has a plurality of attribute values, extracting each of the plurality of attribute values separately by marked identifiers.

7.   The method of claim 1, wherein the step (c) comprises the steps of:

(c-1) converting each sentence of an illustrative corpus into a token string, recognizing dependence relation of an attribute tagging token, generating learning data, and learning the learning data through a predetermined stochastic model; and

(c-2) converting each sentence of the body of the encyclopedia into the token string, recognizing dependence relation of an extraction object tokens, and applying a learning result and the stochastic model to a recognition result, thereby finding and extracting the attribute name and the attribute value of each extraction object token.

8.   The method of claim 7, wherein the step (c-1) comprises the steps of:

(c-1-i) performing morpheme parsing on the illustrative corpus of the encyclopedia, which is tagged with an object name and an attribute, and recognizing a word-phrase unit token string

according to sentences;

(c-1-ii) applying a predetermined dependence rule to a token tagged with an attribute value in the token string, thereby recognizing dependence relation between a governor and a dependent for the object token; and

(c-1-iii) generating the learning data by using the governor and the dependent of each object token as contexts, and storing the learning result in the stochastic model.

9.   The method of claim 7, wherein the step (c-2) comprises the steps of:

(c-2-i) performing the morpheme parsing and object name recognition on the body of the encyclopedia, and converting each sentence into the word-phrase unit token string;

(c-2-ii) designating a token of the token string as an extraction object token, the token of the token string having object name or full morpheme as a noun;

(c-2-iii) applying a predetermined dependence rule to each of the designated extraction object tokens, and recognizing a context token of the governor and the dependent; and

(c-2-iv) applying the extraction object token and the context token to the learning result and the stochastic model, grouping attribute types of the extraction object tokens, and extracting the attribute type of the extraction object tokens

that have highest probabilities with the attribute names and the attribute values of the extraction object tokens.

10. The method of claim 8, wherein, in the dependence rule used to recognize the dependence relation, the governor is a verb phrase nearest to the dependent if the dependent is any one selected from the group consisting of a subjective case, an objective case and an adverbial case, and the governor is a noun nearest to the dependent if the dependent is any one selected from the group consisting of an adnorminal phrase and an adnorminal clause.

11. The method of claim 8, wherein in the dependence rule used to recognize the dependence relation, in case of neighboring nouns and/or object names, a preceding noun or a preceding object name is a dependent and a following noun or a following object name is a governor.

12. The method of claim 8, wherein in the dependence rule used to recognize the dependence relation, when tokens around an object name or nouns are symbols, a verb phrase of a sentence is a governor.

13. The method of claim 7, wherein a maximum entropy model

is used as the stochastic model.

14.　The method of claim 1, wherein the step (d) comprises the steps of:

(d-1) constructing the knowledge base with the attribute name and the attribute values extracted in the step (b); and

(d-2) additionally storing the attribute name and the attribute values extracted as the unstructured information in the knowledge base according to existence of same attribute value of the entry.